NARVA

(6)

DEVELOPMENT OF A SYSTEMATIC METHODOLOGY FOR THE APPLICATION OF JUDGMENTAL DATA TO THE ASSESSMENT OF TRAINING DEVICE CONCEPTS

(10)

MARSHALL A./NARVA, Ph.D.
U.S. ARMY RESEARCH ~~INSTITUTE FOR THE~~ BEHAVIORAL AND SOCIAL SCIENCES
5001 EISENHOWER AVENUE
ALEXANDRIA, VIRGINIA 22333

(12) 15p.

ABSTRACT

Concurrency of development of the training subsystem is rapidly becoming a key concept in the development of new materiel systems. The development of the training subsystem must occur concurrently with that of the prime system, in order to meet the objective of having a total system operational when it is fielded. This requires that the development of the training subsystem commence early in the life cycle of the prime system, i.e., in the conceptual stage. Towards this end a number of activities concerned with concurrent training development have been outlined in such documents as DA Pamphlet 11-25, the Life Cycle System Management Model (13). These activities include the identification of requirements for training devices. Once a decision has been made that a training device is required (and there is an increasing impetus towards utilization of training devices), the training device concept or alternative concepts must be evaluated.

ABSTRACT

Due to the time constraints pushing in from the deadlines associated with the ongoing system development, the training developer must make decisions or judgments early on in the process, when information available from the prime system development may be sketchy. While the development of the training device concept must commence as soon as possible in order to meet the required deadlines, particularly on long lead-time items, "bending of metal" for prototypes cannot begin before the design freeze of the pertinent segment of the prime system. Therefore, assessment of training device concepts becomes more dependent upon the use of analytical as opposed to empirical assessment techniques. Also, the rising costs associated with the development of such devices gives rise to an increased requirement for development of a systematic methodology for the assessment of such concepts.

448 011

One methodological tool, or decision aid, which has been developed to assist in making such judgments is known as TRAINVICE. The original model was developed for the Army Research Institute and was based upon an extensive literature review and analytical work by a team of experienced behavioral scientists (15, 16, 17). A revision of this model based upon experience gained in its utilization and which we may call TRAINVICE II, is being developed with the aims of increasing the practicality and flexibility of its application and making the methodology more amenable to utilization by a wider spectrum of users (9). This paper will concentrate on the rationale for TRAINVICE II and the difficulties involved in the development of a practical model which provides for incorporation of the variable experience and personal judgmental factors of the training developer or analyst vis-a-vis available recommendations or guidance.

TRAINVICE II

Essentially the model provides a framework for asking three questions concerning a training device concept. As given in Figure 1, these are, "What," "Why," and "How." These three questions are asked for each unit of behavior that is being trained by the device. A training device represents a cluster of activities, which become better defined as the concept becomes better defined. This may take the form of definition at the task, subtask, or skill/knowledge level, which represents the component physical and mental activities involved. Such a model as this is best utilized when information is available to the finest meaningful level of detail, as the required judgments can be made more accurately. However, an attempt is being made to have the model accommodate information at a more molar level in order to initiate analysis as early as possible in the development cycle. (A corollary requirement is imposed upon the meteriel developer to make detailed information available as soon as possible to meet the needs of the training developer.)

In order to answer the question of "what" is to be represented in the training device, two judgments are required. The first deals with the requirement for the unit of activity to be represented in the device, while the second ascertains if that unit is covered. These two judgments together have as their objective the determination that the spectrum of activities covered neither exceeds that which is necessary nor leaves out any that should be covered.

The "why" question component deals with a more detailed "defense" of the reason for including the unit of activity which was deemed to be necessary in the previous judgment and which is indeed included, or is to be included, in the device. This assessment is

<u>WHAT</u>

<u>Coverage Requirement</u>

Should this unit of activity be represented in the device?

<u>Coverage</u>

Is this unit of activity represented in the device?

<u>WHY</u>

<u>Training Criticality</u>

What level of proficiency is required in this unit of activity?

<u>Training Difficulty</u>

How difficult will it be to train for this unit of activity?

<u>HOW</u>

<u>Physical Characteristics</u>

How well do the physical characteristics of the device involved in training for this unit of activity meet the perceived requirements and applicable guidelines?

<u>Functional Characteristics</u>

How well do the functional characteristics of the device involved in training for this unit of activity meet the applicable guidelines?

Figure 1. Judgments made in TRAINVICE II

performed relative to two aspects; Training Criticality, which relates to the degree of proficiency required at the end of the training, and Training Difficulty, which considers the degree of difficulty to be expected in training to reach the desired level of proficiency. When these factors are assessed quantitatively, these judgments in essence give a weighting to each required unit of activity covered by the device. It should be noted that the decisions concerning "what" may have to wait upon or be changed by feedback from the decisions concerning "why" an activity unit should be included in the device, depending upon the stage of development of the device concept.

The first two questions dealt with an assessment of the spectrum of activities to be covered by the device, and not with how these activities were to be taught or conveyed by the device; the last question "how" deals with this aspect. The "how" question considers two aspects; do the physical characteristics of the device follow available guidelines of "good practice" and do they fit the perceived requirements of the training device developer, and are these physical characteristics utilized in keeping with available guidelines for "good practice"; are their functional characteristics such as to make for good training?

The assumption is being made that the potential for transfer of training will increase as a function of the degree to which the required activities, physical and mental, are represented in the device and the degree to which the training device follows "good" practice in training in these activities. In addition, each of the activity units is appropriately weighted by the degree of skill required and the degree of difficulty involved.

At this point it may be well to discuss the utilization of the model for both prescriptive and predictive purposes. It is hoped that such a model as TRAINVICE II can be utilized for both purposes, in keeping with the objective of utilization as early as possible in the development cycle. A prescriptive utilization would be for the *a priori* formulation of the training device concept; predictive utilization would be the application of the model to an existing training device concept or prototype to predict the effectiveness of transfer of training which is to be expected through use of the device. When information is incomplete, due to the stage of development of the prime system or the training subsystem, the prescriptive mode takes on added importance. This aspect will be touched upon as we now go through the model in more detail.

Coverage requirement analysis. The first judgment to be made is whether each of the skills (or knowledges) subsumed under the training objective should or should not be included in the training situation. Depending upon the stage of development of the training device, this analysis may help to delineate the range of skills to be represented in the device (the prescriptive mode), as well as assessing the range of skills represented in a device (the predictive mode). The judgments may be recorded simply in the form of a checklist, or, if a quantitative prediction is to be attempted, through the use of 1 or 0, to indicate a requirement for a skill/knowledge, or lack of a requirement, respectively. This analysis is a "gate" only; it determines if the skill should be represented in the training.

Coverage Analysis. The second judgment is made by comparing the skills which have been judged to be required to be included in the training with those that indeed are included in the training device concept. This judgment may not be possible to make in the prescriptive mode, before a training device concept is formulated; however, it could take the form of a checklist to be utilized as the concept takes form, to ensure that all the required skills are included. In the predictive mode, when an existing concept is being assessed, a "1" may be assigned rather that a checkmark, and a "0" utilized, rather than a lack of a checkmark, in order to feed into the overall index of predicted training device effectiveness which will be subsequently derived.

If the coverage requirement rating is "1" and the coverage rating is "0," this would indicate that training in this skill is lacking and steps should be taken to include it or the device will suffer in its overall rating. On the other hand, if the coverage requirement rating is "0" and the coverage rating is "1," this would indicate that unnecessary training is being provided and should be eliminated from the device, or its overall rating will also suffer.

Training Criticality Analysis. For each of the skills that have been judged to be necessary to be represented in the training situation and are indeed represented, a judgment is made as to the degree of proficiency required in that skill at the end of training. In the predictive mode, for the assessment of an existing concept, the following rating scale, adapted from Demaree (3), is used:

1 Should have limited knowledge of subject or skill
2 Should have received complete briefing on subject or skill
3 Should have understanding of subject or skill to be performed
4 Should have complete understanding of subject, or be highly skilled

In making this judgment, the analyst must take into account the nature of the skill and the degree to which the training will be supplemented by subsequent on-the-job training.

This rating could also be used in a prescriptive mode, in making decisions concerning which of the skills is to be represented; in working back to the coverage requirement decision.

Training Difficulty Analysis. In addition to assessing the level of proficiency required for each of the required skills, a judgment is made of the degree of difficulty which is to be expected in attaining that level of proficiency, for the particular skill and trainee population involved. In the predictive mode, the following rating scale, adapted from that of Rankin (11), is used:

1 Minimal or none
2 Some
3 Much
4 Substantial

The various factors which must be considered by the analyst in making this decision are not explicitly delineated or extracted for judgment, but must be integrated into the judgmental process. These factors are the required level of proficiency, the level of skill existing in the trainee, and the inherent level of difficulty of the activity per se. The analyst must rely upon his own experience or knowledge concerning the difficulty involved in training the various skills or rely on others who do have the experience, or attempt to extrapolate from available knowledge concerning related skills.

As was the case for training criticality, this rating may be used in a prescriptive mode, in making decisions concerning which of the skills are to be represented; in working back to the coverage requirement decision.

Physical Characteristics Analysis. Having considered what skills are covered by the training device concept, and why, we now turn to a consideration of how these skills are to be taught. The first judgment in this respect is that concerning the physical characteristics of the device.

A training device may be considered to be a mosaic of specific elements, be they displays, controls, inputs, outputs, or cue-response pairs. These elements may be likened to the simulation elements proposed by Smode (12), with the configuration and operation of the training device determining the spatial placement and temporal sequencing of the elements. As Matheny (8) has pointed out, the assumption may be made that it is perceptual equivalence that results in positive transfer from the training to the operational situation, and it is the adequacy of these simulation elements in terms of perceptual or psychological equivalency, not physical equivalency, which must be judged. The total training device, and indeed each of the elements embedded in that mosaic, may be considered to lie within a space defined by the dimensions of fidelity and abstraction. Levels of abstraction range from the real world to mathematical models representing dynamics taking place in that world, they are analogies of the real world, while fidelity varies as to the comprehensiveness and level of detail to which the external world is represented (5, 10). In order to deal with the multiplicity of degrees of abstraction and fidelity with which training devices may clothe themselves, it was deemed necessary to translate the specific simulation elements into generic characteristics. This brings the number of specific possible forms that each element may assume into manageable proportions and provides for a common base upon which to make the required judgments concerning the stimuli and responses, or physical characteristics of the training device. Such a listing has been taken from that given in the TECEP technique of Braby, et al (2), and is shown as Figure 2.

The analyst must make a judgment as to what generic characteristics are required and rate the existing generic characteristics against this criterion. In order to make this judgment, he must draw upon his own perceived requirements for each of the elements, or turn to others, either directly or through reference to available guidelines, or both. Some guidance may be gleamed from the literature. However, a simplified version of the Aagard and Braby guidelines which is utilized in the ISD model (1, 14) appeared most suitable and has been adapted for use with the TRAINVICE II model to give some guidance to the analyst in making the judgments concerning the physical characteristics, in conjunction with his own perception of the requirements. However, these guidelines were intended for utilization in instructional system development and instructional delivery system or media selection, and not for scrutiny of specific aspects of a training device. The development of more specific guidelines is needed. These guidelines deal for the most part with functional aspects of the training situation, such as the sequencing of learning events. However, selected guidelines may be extracted as being applicable to the design of specific elements of the training device or situation.

The procedure followed in performing the physical characteristics analysis is shown in Figure 3. For most effective utilization of the model, task information down to the skill/knowledge

STIMULUS CAPABILITIES

Visual Form

1. Visual Alphanumeric
2. Visual Pictorial, Plane
3. Visual Line, Plane
4. Visual Object, Solid
5. Visual Environment

Visual Movement

6. Visual Still
7. Visual Limited Movement
8. Visual Full Movement
9. Visual Cyclic Movement

Visual Spectrum

10. Black and White
11. Gray Scale
12. Color

Scale

13. Exact Scale
14. Proportional Scale

Audio

15. Voice Sound Range
16. Full Sound Range
17. Ambient Sounds

Other

18. Tactile Cues
19. Internal Stimulus Motion Cues
20. External Stimulus Motion Cues

TRAINEE RESPONSE MODES

21. Covert Response
22. Multiple Choice
23. Pre-programmed Verbal
24. Free-Style Written
25. Decision Indicator
26. Voice
27. Fine Movement Manipulative
28. Broad Movement Manipulative
29. Tracking
30. Procedural Manipulative

Figure 2. Generic Characteristics List

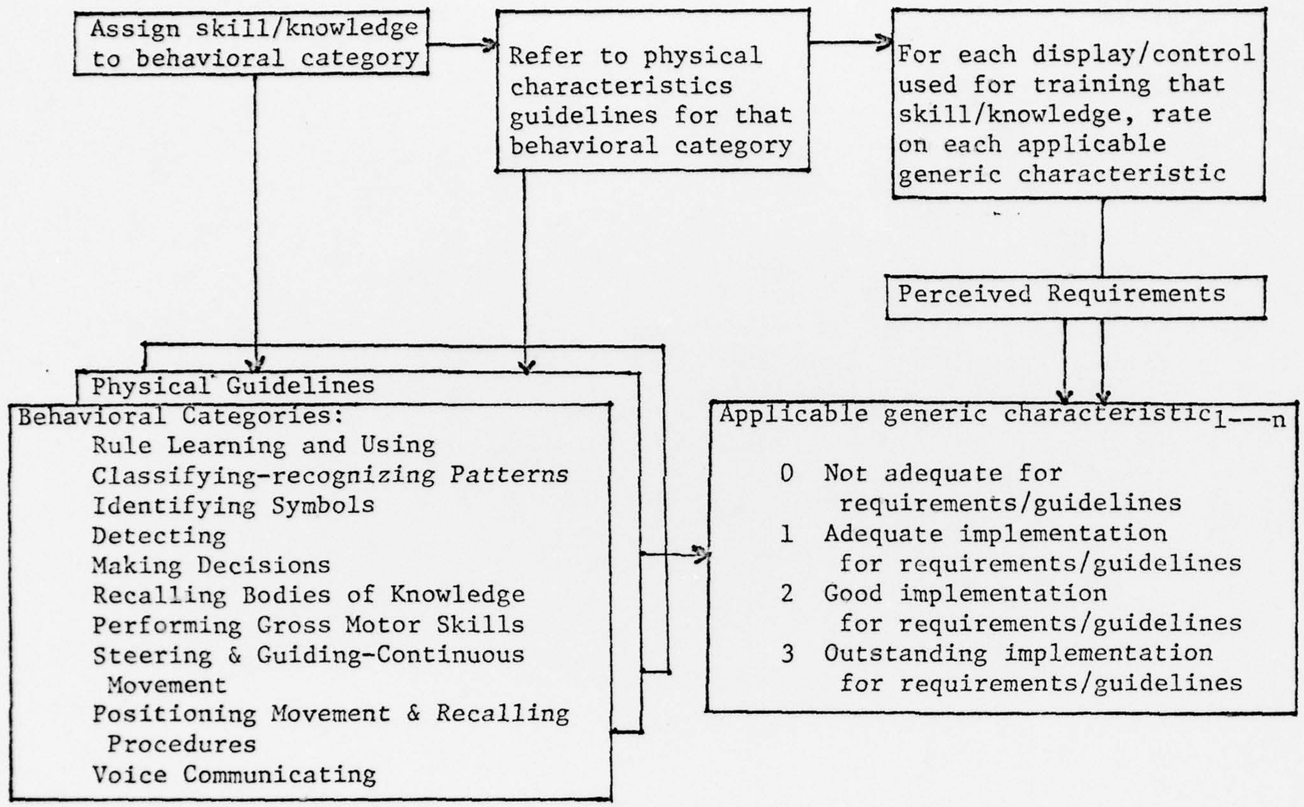

Figure 3. Physical Characteristics Analysis

level should be available. The reason for this is that the guidelines are accessed through means of behavioral categories; there is a different set of guidelines for each of the behavioral categories, and the skill/knowledge involved must be translated into a behavioral category. If an analysis is done at the subtask level, more than one behavioral category may be involved and more guidelines must be considered. Also, for each of the skills, information concerning the displays and controls involved in implementation of that skill should be available, as it is the generic characteristics of these controls and displays which are to be assessed. In the predictive mode, each of the generic characteristics associated with each of the displays or controls (or cues and responses) is rated using the scale shown in Figure 3. Therefore, each of the displays or controls, or physical characteristics, receives a physical characteristics score which is the total of the ratings given on the applicable generic characteristics. In addition, the maximum rating is given for each of the applicable parameters in order to provide a base for assessment of the device in the index subsequently calculated. The pattern of these ratings may serve to "highlight" the various physical characteristics of the device, both those that are outstanding and those that need change. In addition, each of the characteristics may require a justification in order to control for the introduction of unneeded features. In order to derive the Physical Characteristics rating for the skill involved, the ratings given on each of the displays or controls are added to give the total for that skill. Therefore, the presence of a "0" rating does not eliminate that skill from the total rating but does serve to downgrade the total rating for the skill and subsequently for the device.

If information concerning the specific characteristics of the displays or controls involved is not available due to the stage of development, the ratings may fall back to a prescriptive mode, in which the desired generic characteristics, possibly through use as a checklist, are chosen, in lieu of performing an assessment of existing generic characteristics. This is a similar procedure to that utilized in the TECEP technique of Braby et al (2). The TECEP technique gives recommendations as to which delivery system, including generic classes of training devices, permit the application of the learning guidelines for each of the behavioral categories; recommendations based upon the pattern of matching of the generic characteristics inherent in the various delivery systems and those judged necessary by the analyst. Jorgensen (6, 7) has utilized a similar matrix approach in which the generic characteristics judged as being required are matched against various media, of which training devices are one class, in order to select training media most suitable for training various tasks. However, these procedures are intended for the selection of or comparison of various media or instructional delivery systems rather than the scrutiny of a training device per se or comparison of training devices.

Functional Characteristics Analysis. While the physical characteristics analysis is concerned with the analysis of the elements of the training device per se, the functional characteristics analysis is concerned with how these elements are utilized. The operation of the device is compared against guidelines to ascertain to what extent "good" training practices are followed. Once again, these guidelines are those extracted from the ISD model (14).

The procedure followed in performing the functional characteristics analysis is shown in Figure 4. As with the physical characteristics analysis, each of the skills is translated into a behavioral category, which dictates which set of guidelines is to be utilized. The set of guidelines is consulted as the functional, dynamic characteristics of the elements involved in training for that skill are considered. In the predictive mode, each of the guidelines judged to be pertinent to the particular situation involved is rated, using the scale shown in Figure 4. A judgment is made to the extent that that guideline is implemented by the functioning of the cluster of elements used to implement training for that skill. In order to derive a Functional Characteristics rating for the skill, the ratings are added to give the total for that skill. Therefore, the presence of a "0" rating does not eliminate that skill from the total device rating, but does serve to downgrade the rating for the skill, and subsequently of the device, and may also serve to "flag" some aspect that needs to be corrected. In addition the number of applicable guidelines are multiplied by "3," the highest possible rating, to give a baseline against which the derived rating may be compared in the subsequent device index.

If sufficient information is not available to make the required ratings, the procedure may revert to the prescriptive mode, in which the applicable guidelines are selected, and utilized as a checklist for the development of the training device concept. If the analysis must be performed at the subtask level, then more guidelines must be sorted out and consulted, as a larger unit of activity is involved, which may encompass more than one behavioral category.

Derivation of Index of Predicted Training Device Effectiveness. In the predictive utilization of the model, an index is derived. This index follows a procedure discussed by Gagne, Foster and Crowley (4). While not based on one of their formulas directly, it is in keeping with their conclusion that the most useful and

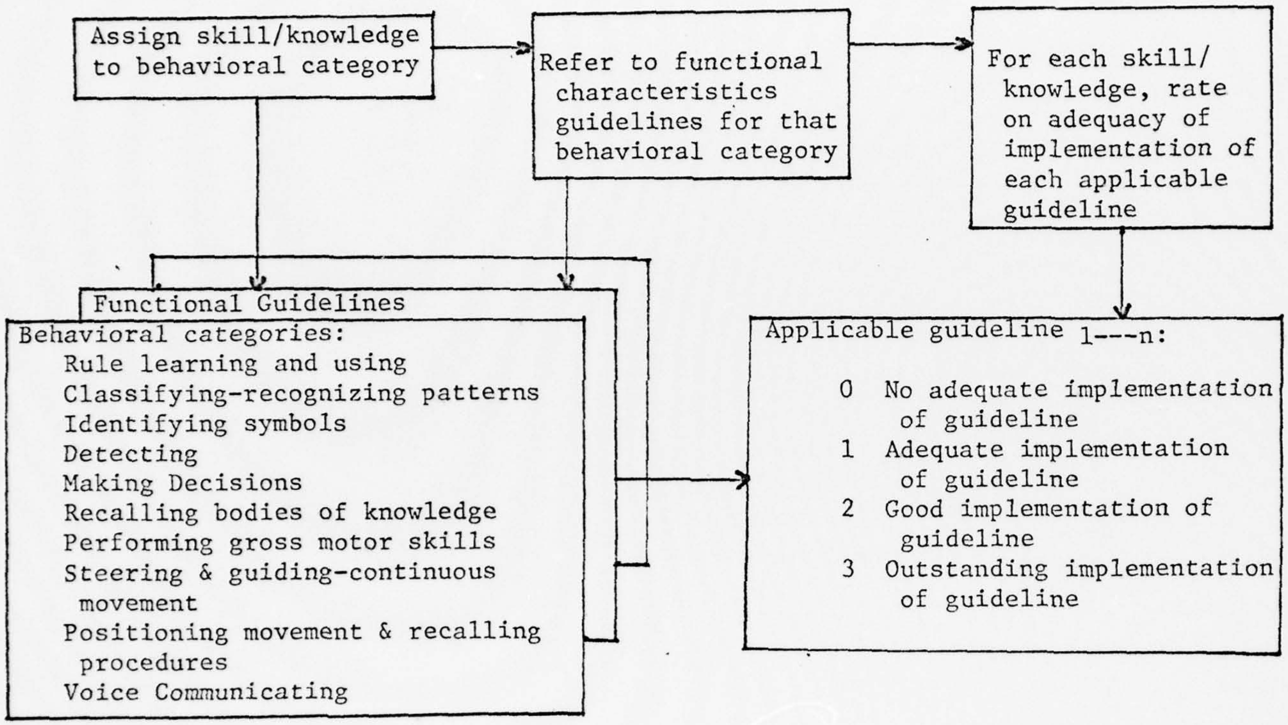

Figure 4. Functional Characteristics Analysis

practical type of formulation is that based on percentage of maximum possible transfer. It assumes that if the device were to follow perfectly all of the guidelines, as judged necessary by the analyst, that maximum transfer, which could be attributed to the device, would be the result. This forms the baseline against which the device under evaluation is compared. This score is weighted by the Coverage Requirement, Coverage, Training Criticality and Training Difficulty scores derived for that skill. The derived score for each skill is compared with the score representing maximum expected transfer. (If a "0" rating is given for either the Coverage Requirement or Coverage Analysis, the total score for that skill is reduced to "0" and makes no contribution to the derived index for the device.) To derive the score for the total device, each of the skill scores is added. Therefore, the index of predicted training device effectiveness is as follows:

$$\frac{\Sigma(CR \times C \times Ci \times D \times (PC + FC))i}{\Sigma(CR \times C \times Ci \times D \times (PC_{max} + FC_{max}))i}$$

where:
- CR Coverage Requirement Score
- C Coverage Score
- Ci Training Criticality Score
- D Training Difficulty Score
- PC Physical Characteristics Score
- FC Functional Characteristics Score
- PC_{max} Maximum Possible Physical Characteristics Score
- FC_{max} Maximum Possible Functional Characteristics Score for each skill.

This equation will yield an index ranging from 0 to 1. The larger the index, the larger the number of required skills represented and the higher were the ratings given on the Device Characteristics Analyses; therefore, presumably the greater the potential for transfer of training.

Conclusions

(1) Due to the increased emphasis on the concurrent development of training subsystems, the increased impetus towards the utilization of training devices, and the rising costs associated with the development of such devices, there is a need for the development of a systematic analytical methodology for the assessment of training device concepts which may be applied early in the development cycle of the materiel system.

(2) Such a methodology must permit the integration of the variable judgmental processes of the training developer and available guidance.

(3) A model has been described which provides a framework for the application of judgmental data concerning aspects of the training situation which have been hypothesized as having an impact upon the effectiveness of transfer from the training to the operational situation. Judgments are made concerning coverage of required skills, the weighting to be assigned to these skills from the aspects of training criticality and difficulty, and the effectiveness with which the physical and functional characteristics of the training situation follow guidelines of good practice.

(4) In order to apply such a methodology as early as possible in the development cycle, a prescriptive mode of utilization, concerned with assisting in the formulation of training device concepts, is provided for.

(5) There is a need for the development of more specific guidelines which may be applied to the assessment of training device concepts.

References

1. Aagard, A., and Braby, R. Learning guidelines and algorithms for types of training objectives. TAEG Report No. 23, U.S. Navy, Training Analysis and Evaluation Group, Orlando, FL, March 1976.
2. Braby, R., Henry, J., Parrish, W., and Swope, W. A technique for choosing cost-effective instructional delivery systems. TAEG Report No. 16, U.S. Navy, Training Analysis and Evaluation Group, Orlando, FL, April 1975.
3. Demaree, R. Development of training equipment planning information. ASD Technical Report 61-533, U.S. Air Force, Aerospace Medical Laboratory, Wright-Patterson Air Force Base, OH, October 1961.
4. Gagne, R., Foster, H., and Crowley, M. The measurement of transfer of training. Psychol. Bulletin, 1948, 45, 97-130.
5. Haythorn, W. W. Information systems simulation and modeling. Paper presented at First Congress on the Information System Sciences, Hot Springs, VA, November 19, 1962. (cited in Obermayer (10)).
6. Jorgensen, C. A methodology and analysis for cost-effective training in the AN/TSQ-73 Missile Minder. Army Research

Institute for the Behavioral and Social Sciences, Ft Bliss Field Unit, El Paso, TX, September 28, 1976.

7. Jorgensen, C. A method for cost training effectiveness analysis. Paper presented at Military Operations Research Symposium, Monterey, CA, 13-15 December 1977.
8. Matheny, G. Training simulator characteristics: Research problems, methods, and performance measurements. Proceedings, Aircrew Performance in Army Aviation, U.S. Army, Office of the Chief of Research, Development and Acquisition, July 1974.
9. Narva, M. Formative utilization of a model for the prediction of the effectiveness of training devices. Paper presented at Army Operations Research Symposium, Ft Lee, VA, 12 October 1977.
10. Obermayer, R. Simulation, models, and games: Sources of Measurement. Human Factors, 1964, 6, 607-619.
11. Rankin, W. Task description and analysis for training system design. Technical Memo 74-2, U.S. Navy, Training Analysis and Evaluation Group, Orlando, FL, January 1975.
12. Smode, A. Human factors inputs to the training device design process. Technical Report NAVTRADEVCEN 69-C-0298-1, Naval Training Device Center, Orlando, FL, September 1971.
13. U.S. Army, Life cycle system management model for Army Systems, DA Pam 11-25, May 1975.
14. U.S. Army, Interservice procedures for instructional system development. TRADOC Pam 350-30, U.S. Army Training and Doctrine Command, Ft Monroe, VA, August 1975.
15. Wheaton, G., Rose, A., Fingerman, P., Korotkin, A., and Holding, D. Evaluation of the effectiveness of training devices: Literature review and preliminary model. Research Memorandum 76-6, U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, April 1976a.
16. Wheaton, G., Fingerman, P., Rose, A., and Leonard, R. Evaluation of the effectiveness of training devices: Elaboration and application of the predictive model. Research Memorandum 76-16, U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, July 1976b.
17. Wheaton, G., Rose, A., Fingerman, P., Leonard, R., and Boycan, G. Evaluation of the effectiveness of training devices: Validation of the predictive model. ARI Technical Report TR-76-A2, U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA, October 1976c.